Mol Gen Genomics (2004) 271: 752–760
DOI 10.1007/s00438-004-1031-7

ORIGINAL PAPER

C. L. Bassett · M. L. Nickerson · R. E. Farrell Jr.
M. Harrison

Multiple transcripts of a gene for a leucine-rich repeat receptor kinase from morning glory (*Ipomoea nil*) originate from different TATA boxes in a tissue-specific manner

Received: 4 November 2003 / Accepted: 21 May 2004 / Published online: 19 June 2004

Abstract TATA boxes are the most common regulatory elements found in the promoters of eukaryotic genes because they are associated with basal transcription initiation by RNA polymerase II. Often only a single TATA element is found in a given promoter, and tissue-, stage- and/or stimulus-specific expression occurs because the TATA box is associated with other *cis* -acting elements that enhance or repress transcription. We used software tools for gene analysis to assist in locating potential TATA box(es) in an AT-rich region of the promoter of a gene, *inrpk1*, which codes for a leucine-rich receptor protein kinase in morning glory (*Ipomoea nil*). Through the use of RT-PCR and various combinations of forward primers bracketing most of the promoter region we were able to define the 5′-ends of transcripts in this region. The region was then targeted for analysis by RNA Ligase-Mediated-5′ Rapid Amplification of cDNA Ends (RLM-5′ RACE) to identify the transcript initiation site(s). Positioning of initiation sites with respect to TATA boxes identified by gene analysis tools allowed us to identify three operational TATA elements which regulate basal transcription from this gene. Two TATA boxes were responsible for all of the *inrpk1* transcripts found in leaves and cotyledons, and about 25–30% of the transcripts in roots. A third TATA box was involved only in expression in roots and accounted for the remaining 50–70% of root transcripts. RNAs expressed from this element lack two potentially functional upstream AUG codons, and may be translated more efficiently than transcripts originating from the other TATA boxes.

Keywords Transcription initiation · TATA element · Upstream open reading frames (uORFs) · RNA Ligase-Mediated (RLM)-5′ RACE · Translational control

Communicated by D. Y. Thomas

C. L. Bassett (✉)
The Appalachian Fruit Research Station,
USDA-ARS, 2217 Wiltshire Road,
Kearneysville, WV 25430, USA
E-mail: cbassett@afrs.ars.usda.gov
Fax: +1-304-7282340

M. L. Nickerson
Laboratory of Immunobiology,
National Cancer Institute-Frederick,
Bldg. 560, Room 12-69, Frederick,
MD 21702, USA

R. E. Farrell Jr.
Department of Biology,
Pennsylvania State University,
1031 Edgecomb Avenue, York,
PA 17403, USA

M. Harrison
Biological Sciences, Marshall University,
1 John Marshall Drive, Huntington,
WV 25755, USA

Introduction

Regulation of gene transcription in eukaryotes is accomplished at multiple levels, thereby facilitating fine-tuning of the expression of individual genes. Since transcript initiation is the first step in gene expression, it is commonly observed to be a primary target of control. Although enhancers and silencers are known to affect the quantitative regulation of gene expression, the primary importance of sequences upstream of the coding region (i.e. the core promoter and 5′ leader sequences) for the control of basal transcription has been thoroughly documented (reviewed by Lewin 2000; Hochheimer and Tjian 2003).

In eukaryotes, the basic transcriptional control element is the TATA box, but in animal promoters three additional elements are also associated with transcript initiation. Two of these elements, BRE (Transcription Factor IIB [TFIIB] recognition element; Lagrange et al. 1998) and DPE (downstream promoter element; Burke and Kadonaga 1996), are found most often in promoters that lack a TATA element. The third element, the initiator (Inr), comprises sequences surrounding the site of transcription initiation and can operate alone or in

conjunction with other elements, particularly the TATA box (Smale and Baltimore 1989). In plants, only the TATA elements have been described so far (Joshi 1987).

TATA boxes bind the TATA-binding protein and associated factors making up TFIID (Transcription Factor IID; Patikoglou et al. 1999). The binding of TFIID permits assembly of RNA polymerase II (RNA Pol II) and associated general transcription factors into a pre-initiation complex which directs basal gene expression (Pan et al. 2000; reviewed in Wassarman and Sauer 2001; Burley and Kamada 2002). Recent crystallographic analysis of the yeast RNA Pol II provides a model in which the duplex DNA follows a straight path through a cleft that extends from one side of the enzyme to the other (Cramer et al. 2001). The proposed crystal complex suggests that a spacing of approximately 25 bp between the TATA box and the transcription start site (TSS) most easily fits into the Pol II active center; this distance is consistent with the observation that the majority of genes initiate transcription 30–60 nts (nucleotides) downstream of the TATA box (reviewed by Shenk 1981; Joshi 1987; Bucher 1990).

TATA sequences are also functionally linked to other regulatory *cis* -acting elements, such as GC-, CAAT-, and G-boxes (Shenk 1981; Bucher 1990; Meshi and Iwabuchi 1995), and it is this linkage that provides tissue or organ specificity, and differential expression in response to chemical and physical signals. Therefore, to understand gene regulation it is important to identify the TATA box, the TSS and other *cis* -acting regulatory elements known or suspected to be associated with the control of gene expression. Although the identification of consensus TATA boxes is not especially difficult using available gene analysis software, in some cases these elements cannot be identified with confidence, particularly when the region 5′ of the coding sequence is AT-rich and contains both consensus and/or non-consensus TATA elements. To understand the basis for differences in expression previously reported for a leucine-rich receptor kinase gene, *inrpk1*, from morning glory (Bassett et al. 2000), we have undertaken an analysis of the 5′ upstream region of this gene using a variety of promoter search programs to identify potential TATA boxes. The presence of three operational TATA elements was confirmed using RLM-5′ RACE to identify the TSSs. Analysis of RNAs expressed in different tissues indicated that one TATA element was root-specific. The implications of the use of this element exclusively in roots for the regulation of expression of the putative polypeptide product of *inrpk1* (INRPK1) are discussed in light of what is known about multiple transcript initiation sites and eukaryotic translation control.

Materials and methods

Tools used for *in silico* promoter analysis

Initial searches for promoter sequence analysis tools on the internet led to the identification of two primary sites: The Baylor College of Medicine (BCM) Search Launcher (at http://searchlauncher.bcm.tmc.edu/) and L'Atelier BioInformatique de Marseille (ABIM) On Line Analysis Tools (http://www.up.univ-mrs.fr/~wabim/english/logligne.html). Other useful sites included: BioInformatics and Molecular Analysis Section (BIMAS, at http://bimas.cit.nih.gov/) and BioBencHelper (http://biobenchelper.hypermart.net). The promoter search programs used in this study were accessed through one or more of the sites listed above and are summarized in Table 1.

Table 1 Results of *in silico* analysis of the 1200-bp sequence immediately upstream of the translation start site in the *inrpk1* gene

Tool used	Source	Analysis method	Database used	Number of TATA boxes found	Position(s)[a]	Reference
Signal Scan	BIMAS	Weight matrix	TRANSFAC/ TFD/IMD	1	533[b]	Prestridge (1991)
NNPP	ABIM/BCM	Neural network	–[c]	1	557	Reese et al. (1996)
HCtata	ABIM	Neural network	EPD	2	557[d], 683	Milanesi et al. (1996)
EUKPROM	PC/GENE	Weight matrix	EPD	2	237, 683	Bucher (1990)
TESS	ABIM/BCM	String[e]	TRANSFAC	7	237, 801[f], 952, 1013[g], 1034, 1067, 1083, 1188	Schug and Overton (1997)
MatInspector	ABIM/BCM	Weight matrix[h]	TRANSFAC	8	25, 193, 237, 524[i], 533, 557, 659, 683	Quandt et al. (1995)
TFSEARCH	ABIM	Weight matrix[h]	TRANSFAC	3	25, 237, 524	Akiyama (1995)
FunSiteP	ABIM	Weight matrix	TRANSFAC	2	798, 893[j]	Kondrakhin et al. (1995)

[a]Positional accuracy is ±3 nt
[b]Position at 533 was obtained using IMD (bird); no TATA boxes were found with TRANSFAC or TFD
[c]Database not specified; position at 557 was obtained with a cutoff value of 0.5
[d]Scored as 'weak'
[e] 20% mismatch allowed
[f] 801 is an imperfect repeat of the direct repeats at positions 952 and 1067
[g]The region at position 1188 is a direct repeat of 1013 and 1083
[h]Core similarity was set to 70; matrix similarity was set to 75 (MatInspector); threshold = 75.0 (TFSEARCH)
[i]The TATA element beginning at nt. 524 overlaps the one beginning at nt 533
[j]Positions at nt 798 and nt 893 represent potential transcript initiation sites with a weight value of 0.5

RT-PCR analysis

inrpk1 (GenBank Accession No. U77888) codes for a leucine-rich repeat receptor protein kinase, and was subcloned from a genomic library prepared from morning glory [*Ipomoea* (*Pharbitis*) *nil* Roth, cv. Violet] DNA using a fragment from the maize receptor kinase gene pZM1cat (Walker and Zhang 1990) as a probe (Bassett et al. 2000). Additional upstream sequence information was obtained by gene-specific PCR extension from the library clone (described in Bassett et al. 2000) and included approximately 1144 bp upstream of the putative translation initiation site. Total RNA was isolated according to the procedure of Short and Torrey (1972) from cotyledons, first leaves and roots obtained from plants sampled 5 and 6 days after short-day induction of flowering (induced) or continuous exposure to light (vegetative).

cDNA was prepared from total morning glory RNAs using the Advantage RT for PCR kit (Clontech, Palo Alto, Calif.) and a combination of random hexamers and oligo(dT) to prime cDNA synthesis (Bassett et al. 2000). The primers used for subsequent PCR amplification were: MG389F (5′-GCGTCGGGCCGTTG TAGTGTG-3′), MG483F (5′-TCTGTCGCGACAGA-GACCGCGT-3′), MG897F (5′ CAATAGTTGAATC TTTTGCAGATGC-3′), 5′1136L (5′-CTGACTACC-CACCCAGAGGGTGC-3′), MG1243L (5′-GTGA-GAGCAGAGCTGCTCCATCT-3′). The following touchdown PCR cycling profile was applied to 20-µl reactions: 94°C for 2.5 min, then 15 cycles of 94°C for 1 min, 70°C for 1 min, 72°C for 2.5 min, with a decrease of 0.5°C in annealing temperature per cycle, followed by 30 cycles of 94°C for 1 min, 63°C for 1 min, 72°C for 2.5 min, with a final extension time of 15 min at 72°C.

Following amplification a 10-µl aliquot was removed from each reaction, electrophoresed in a 1% agarose gel in TAE buffer (40 mM TRIS-acetate pH 8.5, 2 mM EDTA) and visualized with SYBRGold (Molecular Probes, Eugene, Ore.) stain. Band images were captured on a STORM 860 fluorescence imager (Molecular Dynamics, Sunnyvale, Calif.).

RLM-5′ RACE analysis of *inrpk1*

RLM-5′ RACE was performed using the RLM-RACE system (Ambion, Austin, Tex.) according to the manufacturer's protocol. Briefly, calf intestinal phosphatase was incubated with total RNA to remove 5′ phosphate groups from degraded mRNAs, stable uncapped RNAs and any contaminating DNA. Subsequently, 7-methylGDP was removed from capped mRNAs (assumed to be primary transcripts) using tobacco acid pyrophosphatase. The 5′ RACE adapter was ligated to the decapped mRNAs with T4 RNA ligase. The resulting RNAs were reverse transcribed into cDNA with MMLV reverse transcriptase primed by random hexamers. PCR was performed on the cDNAs using the gene-specific primer, 5′1136L, and the 5′ RACE outer primer supplied with the kit. A second round of PCR was performed on the products from the first PCR using a nested gene-specific primer, MG836R (5′-GAATGTAAGGAACC AACACCAAGACTC-3′), coupled with the 5′ RACE inner primer supplied with the kit. Typical PCR conditions were: 94°C for 3 min, then 35 cycles of 94°C for 30 s, 60°C (annealing temperature) for 30 s, followed by 72°C for 2 min, with a final extension at 72°C for 7 min. Products of the PCRs were separated in 3% agarose/TAE gels and visualized by ethidium bromide staining. The PCR products were then TA cloned into pCR2.1 (InVitrogen, Carlsbad, Calif.), and recombinant inserts from 50–100 independent transformants from each tissue/treatment sample were analyzed by colony PCR. Positive colonies were then grown in selective medium, and the recombinant plasmids were isolated using the QIAprep Spin miniprep kit (Qiagen, Valencia, Calif.). For each tissue/treatment 10–20 independent clones were sequenced with an ABI Big Dye v. 3.0 DNA sequencing kit (Applied Biosystems, Foster City, Calif.) to confirm their identity and localize the 5′ ends on the *inrpk1* genomic sequence. Because of size differences among the clones it was possible to refer the sequence information back to the colony PCR results and thus to correlate the sequence information with the sizes of the inserts. In this way we were able to obtain information for a total of 212 clones (26–61 independent clones per tissue/treatment).

Intron analysis

RT-PCR was performed on cDNAs synthesized from total RNA as described above. The primers used in the analyses were MG598F (5′-CTTCCTTCACCGCAT-CAAGCC-3′) paired with 5′1136L, and XLR6L (5′-CCCTATTATTCCTTCCCCCTTCTG-3′) paired with MG1243L. Reaction conditions were identical to those for RT-PCR analysis, and the products were separated and documented as described above.

Results

The first 1260 bp of *inrpk1* (ATG_{1144} = putative translation start codon) contain a number of AT-rich regions (Fig. 1), which were analyzed *in silico* for the presence of TATA element(s) using twelve different programs. TSSG, TSSW, PromoterInspector, TESS weight matrix and Proscan were unable to assign a consensus TATA box to any of the AT-rich regions (data not shown). In contrast, other programs (Table 1) found several putative TATA boxes, ranging from two (HCtata, EUKPROM and FunSiteP) to as many as eight (TESS String Search and MatInspector). Signal Scan and NNPP each found a single TATA match at positions 533 and 557, respectively. The locations of putative TATA elements

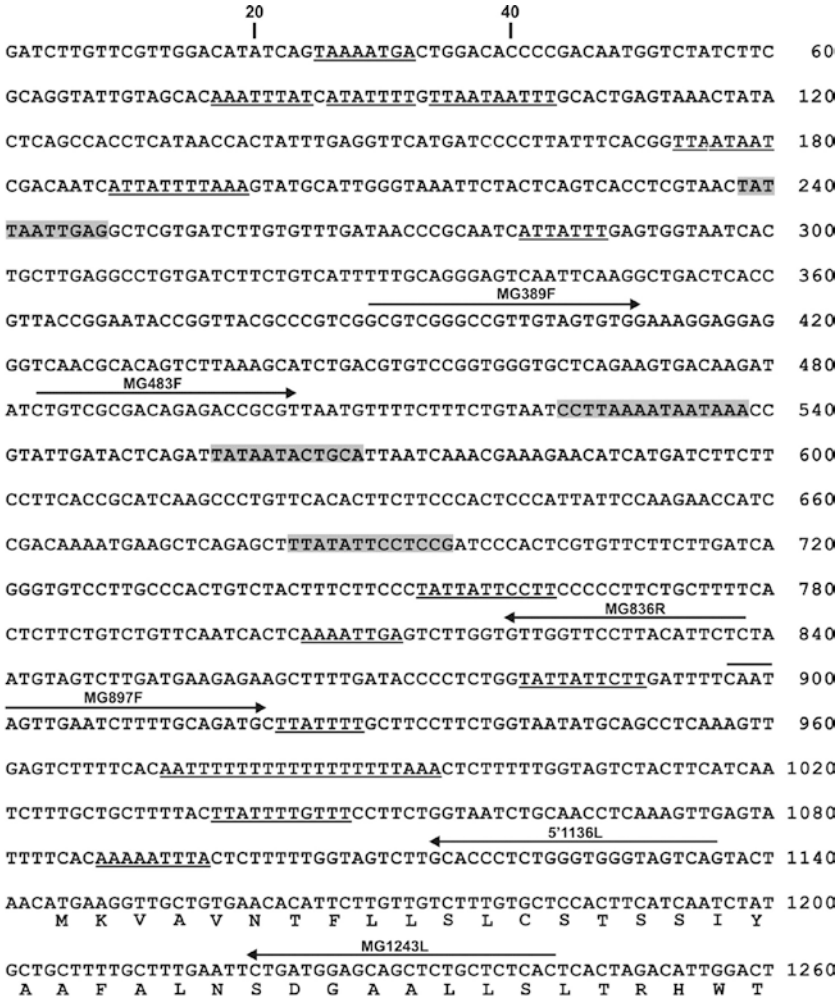

Fig. 1 Sequence of the region upstream of the putative translation start in *inrpk1*. The first 1260 bases of *inrpk1* are shown with AT-rich segments *underlined*. Four possible TATA boxes identified by various software programs are *shaded*. Forward and reverse *arrows above* the sequences represent forward and reverse primers used in the preliminary analysis of the promoter region to identify potential TSSs. The putative translation start site at nucleotide 1144 is indicated by conceptual translation of the polypeptide

found by three or more programs are shown in Fig. 1 (shaded sequences).

Using the information obtained from the promoter analysis programs, we designed primers to ascertain whether or not transcripts encoding *inrpk1* extended upstream beyond nt 389. Three primer combinations, MG389F/5′1136L, MG483F/5′1136L and MG897F/MG1243L, were used to amplify cDNA synthesized from RNA isolated from cotyledons, first leaves and roots from plants that had been induced to flower. Products of the predicted sizes were obtained with all the primer pairs from a positive control genomic DNA template (λ2C, Fig. 2), but none was generated when either MG389F/5′1136L or MG483F/5′1136L was used to amplify the cDNAs (Fig. 2, lanes PP1 and PP2). Similar results were obtained using cDNAs prepared from the same tissues sampled from vegetative plants (data not shown). In contrast, a band of the predicted size was obtained in PCRs primed by MG897F/MG1243L, the primer pair located closest (only 247 nt upstream) to the putative translation start site for the full length polypeptide (Fig. 2, lanes PP3). This product was also generated from cDNAs recovered from vegetative plants (not shown). The result of this experiment defined the 5′ limits of transcripts of *inrpk1*, indicating that no functional TSS lies upstream of nt 483, at least in the tissues examined. We therefore eliminated those TATA elements predicted at positions 25 (not shown) and 237 from further consideration. This information also indicated that the full-length *inrpk1* transcript initiated upstream of nt 897 (5′ end of MG897F), which was consistent with the possibility that the predicted TATA boxes identified at positions 524, 557 and 683 might be functional.

In order to determine which of the potential TATA elements was in fact functional, we determined the location of the transcript start site in cotyledons, first leaves and roots using RLM-5′ RACE. A reverse primer beginning at nt 836 (MG836R, Fig. 1) was used to analyze cDNAs synthesized from de-capped mRNA templates as described in Materials and Methods. Products from these reactions were cloned and sequenced to identify the 5′ end of the mRNA. At least 26 independent clones were analyzed for each tissue examined. The results are shown in Fig. 3 and summarized in Table 2. The majority of transcripts from first

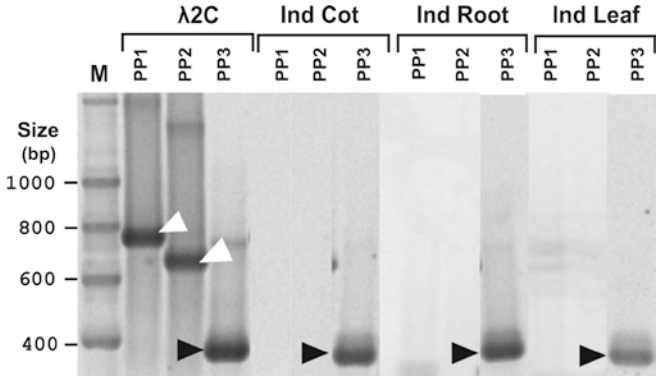

Fig. 2 Preliminary analysis of the *inrpk1* promoter region for potential transcript initiation sites. Total RNA was isolated from cotyledons, roots and first leaves of plants induced to flower or left in a vegetative state. cDNAs were prepared as described in Materials and Methods, and used as templates for touchdown RT-PCR. Aliquots (10 μl) of PCR products were fractionated by electrophoresis in 1% agarose gels and visualized with SYBRgold stain. Lanes PP1, products obtained with primer pair MG389F/5′1136L; PP2, primer pair MG483F/5′1136L; PP3, primer pair MG897F/MG1243L; M, molecular size markers. λ2C is a genomic clone serving as positive control for PCR products; Ind, induced; Cot, cotyledon. The *arrowheads* indicate predicted products of the three sets of primer pairs: *upper open arrowhead*, 747-bp product; *lower open arrowhead*, 653-bp product; *filled arrowheads*, 346-bp product

leaves of induced and vegetative plants initiated at positions 605 and 590, respectively. A smaller, but significant percentage of transcripts also initiated at nt 590 in induced first leaf and at nt 643 in vegetative first leaf. Only 4% of transcripts from both classes of leaves started at nt 561. The vast majority of transcripts from induced cotyledons initiated at position 600, although, as seen with leaves, a few (3%) initiated at nt 561.

In contrast to the results with leaves and cotyledons, the majority of transcripts from induced and vegetative roots initiated much further downstream, starting at nt 716. A significant percentage also initiated at position 788 (31%) in induced roots, while 6–8% initiated at positions 561, 590 and 637. In vegetative roots 16% of transcripts initiated at nt 625, and even fewer initiated at nt 590 and nt 600 (7%). No transcripts initiating at nt 561 were detected.

Because the 5′ leader sequences of *inrpk1* mRNAs are long compared to those found in other plant leaders, and because there are six potential translation start codons upstream of the start site at nt 1144 (Fig. 3), we examined the *inrpk1* 5′ leaders for the presence of introns. Two sets of primers were used to analyze this region. MG598F and MG1243L were designed to target transcripts originating between nt 561 and nt 605, and XLR6L and MG1243L were designed to target transcripts originating at or beyond nt 716. The results are shown in Fig. 4. Neither primer set detected smaller transcripts in the cDNAs derived from total RNA, indicating that no introns larger than 50 bp were present, which is the limit of resolution in a 1% agarose gel.

Discussion

Studies using a variety of animal models indicate that multiple promoters for a given gene are not uncommon (for example see Zhang et al. 1997; Fra et al. 2000). However, with the exception of genes encoded by mitochondrial (Lupold et al. 1999) and chloroplast genomes (Haley and Bogorad 1990), there are few examples of alternative promoters in the nuclear genomes of plants. In some cases the alternative promoter is found in the first intron, resulting in an mRNA that skips the first exon or encodes a truncated polypeptide (Sheen 1991; Tamaoki et al. 1995). Different TATA boxes in the same promoter of a carrot gene encoding dihydrofolate reductase-thymidylate synthase specify polypeptide isoforms that are targeted to the cytoplasm or the chloroplast (Luo et al. 1997). Similarly, Mireau et al. (1996) detected multiple transcripts encoding both cytosolic and mitochondrial alanyl-tRNA synthetases apparently arising from a single TATA element. However, these examples involve either two separate promoters (Sheen 1991; Tamaoki et al. 1995) or a single TATA element (Mireau et al. 1996).

A situation similar to the multiple TATA elements in a single promoter that we observe with *inrpk1* is seen in tomato. Two different TATA elements direct transcription of long and short mRNAs from a phenylalanine ammonia-lyase gene (*PAL5*) and a Ca^{2+}-ATPase gene (*LCA1*) promoter (Lee et al. 1994; Navarro-Avino et al. 1999). The TATA elements in *PAL5* respond differently to various environmental stimuli, and therefore provide additional flexibility in regulating the expression of this *PAL* gene in response to stress. In the case of *LCA1*, the long transcript has a 5′ leader sequence that is >900 bp longer than the 5′ leader of the short mRNA. Interestingly, this leader has numerous upstream AUGs which could potentially affect translation efficiency (see below) if not removed by splicing.

Although sequences lying upstream of the putative translation start site of *inprk1* contain a number of AT-rich regions, only four of these were recognized as TATA elements by various promoter analysis programs. MatInspector was the only program that identified all four TATA boxes. Confirmation that some of these TATA elements were functional was obtained with the RLM-5′ RACE technique, which identified TSSs within 30–40 bp of the center of the TATA boxes.

The TATA element beginning at nt 524 deviates from the consensus TATA box defined by Bucher (1990), but is similar to an artificial TATA box construct that results in low RNA polymerase II activity in vitro in the absence of other *cis* -acting elements (Wang et al. 1996) and is also similar to the TATA box associated with the petunia *cab22R* gene (Gidoni et al. 1989). The low activity of these TATA boxes suggests a plausible reason why only a small number of transcripts initiate downstream of $TATA_{524}$ at nt 561. The sequence context around the TSS at nt 561 is an exact match to the

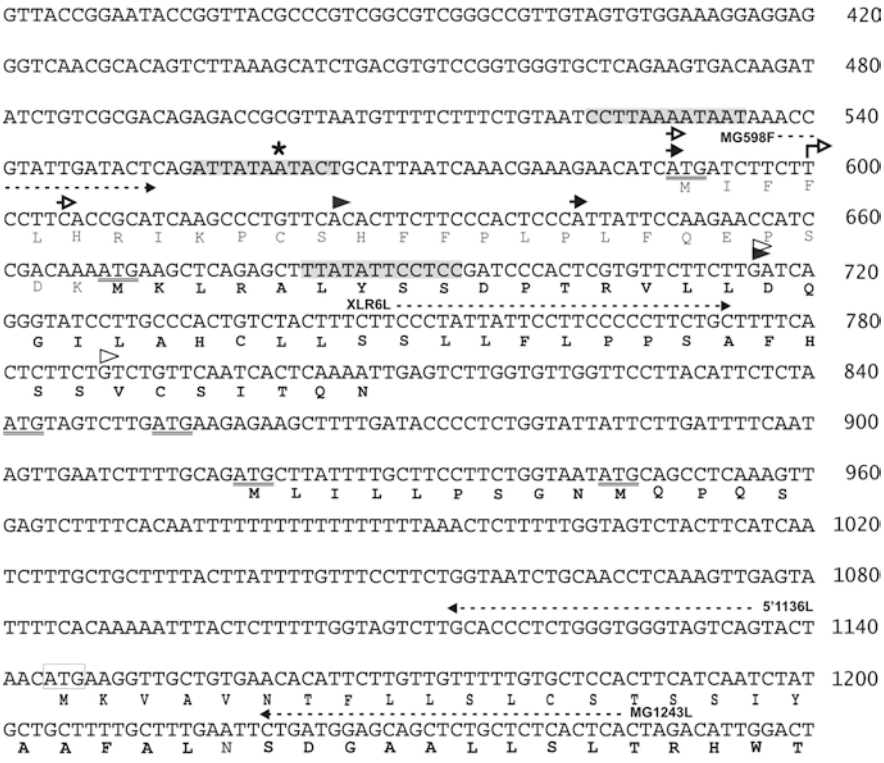

Fig. 3 Promoter region of *inrpk1*. Clones obtained from RLM-5′RACE experiments were isolated and sequenced as indicated in Materials and Methods. The putative translational start site at nucleotide position 1144 is *boxed*, and the translated amino acids appear *beneath* their codons. The *arrows* mark the positions of TSSs identified by RLM-5′RACE. The *angled arrow* (nt 600): major TSS in induced cotyledons; *open arrows*: (nt 590 and 605): major TSSs for induced first leaf; *closed arrows*: major TSSs in vegetative first leaf (nt 590 and 643); *open and closed arrowheads*: major TSSs in induced (nts 716 and 788) and vegetative (nts 625 and 716) roots, respectively. The *asterisk* (nt 561) indicates a minor TSS seen in all tissues except vegetative roots. The *shaded* sequences represent TATA boxes. Six uAUG codons upstream of nt 1144 are *double underlined*. Two in-frame uORFs are identified by the cognate amino acids *below* the sequence. The *dotted arrows* indicate positions of primers used in the intron analyses

TSS of the soybean nodulin gene 23 (Mauro et al. 1985), and may aid in determining transcript initiation from this site. It is interesting to note that, unlike their animal counterparts, few (less than 15% of) plant genes initiate transcription at consensus (PyPyPyCANTPyPyPyPy) cap sites (Joshi 1987).

The TATA element at nt 557 is most likely to provide basal control for transcripts initiating at positions 590, 600 and 605. These transcripts are found in induced and vegetative first leaves, as well as in induced cotyledons, and represent the major mRNAs synthesized in these organs under the conditions examined. Interestingly, transcripts initiating at nt 561 and nt 590 contain six upstream AUGs (uAUG) encoding three uORFs >12 codons long (Fig. 3). uORF1 (73 codons) begins with $uAUG_{590}$ and uORF2 (47 codons) begins in-frame with ORF1 at nt 668. uORF5 begins at nt 918 and potentially encodes a peptide of 14 amino acids. Like the AUG at nt 1144, the AUGs of uORFs 1 and 2 closely match the consensus AUG sites reported by Kozak (1987) and Joshi et al. (1997), suggesting that they could initiate translation. Transcripts beginning at nt 600 and nt 605 lack uORF1, but retain uORF2. mRNA leaders with such long uORFs have been shown to re-initiate translation of downstream ORFs inefficiently (Kozak 2001). uAUG3 and uAUG5 are found in the poorest context for efficient initiation of translation. However, several plant genes with long 5′ leaders containing uAUGs that differ from the consensus (weak initiation codons) have been shown to be translated reasonably

Table 2 Percentage of *inrpk1* transcripts initiating at specific nucleotides in different tissues from induced and vegetative plants

Tissue	TSS (nucleotide position[a])									
	Number of clones analyzed	561	590	600	605	625	637	643	716	788
Induced cotyledons	32	3%	-[2]	97%	-	-	-	-	-	-
Induced first leaves	26	4%	19%	-	77%	-	-	-	-	-
Vegetative first leaves	49	4%	71%	-	-	-	-	25%	-	-
Induced roots	51	6%	8%	-	-	-	8%	-	47%	31%
Vegetative roots	61	-	7%	7%	-	16%	-	-	70%	-

[a]TSS is within 3 bp
[b]The *dashes* indicate that no transcripts originating from that position were detected

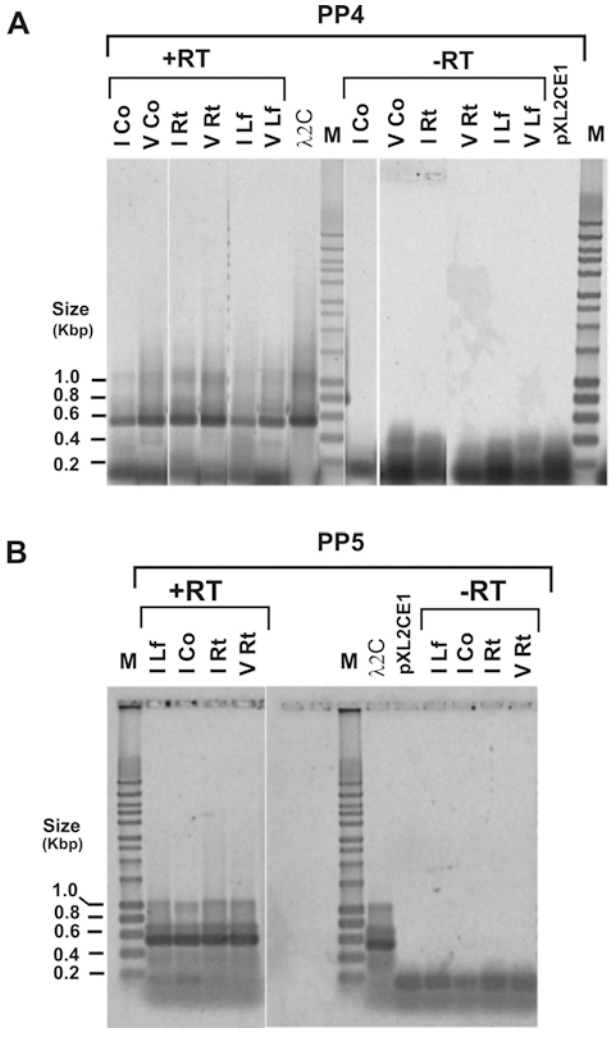

Fig. 4A, B Analysis of the *inrpk1* promoter region for the presence of introns. Primers were designed to detect the presence of introns in the promoter region between positions 598 and 1243, and between 750 and 1243 (see also Fig. 3). PCRs were carried out and the products were separated as described in Materials and Methods. Products were visualized using SYBRgold stain. **A** PCRs were performed on cDNAs from induced (I) and vegetative (V) cotyledons (Co), roots (Rt) and first leaves (Lf). λ2C is a genomic clone, which serves as a positive template control; 2XLCE1 is a genomic clone used as a negative control. RT, reverse transcriptase; PP4, primer pair MG598F and 5′1136L. **B** PCRs were primed with PP5 (primers XLR6L and MG1243)

well both in vivo and in vitro (Michelet et al. 1994; Procissi et al. 2002).

Interestingly, the two major root transcripts initiate downstream of uORFs 1 and 2, and are probably controlled by the TATA element at nt 683. This observation is reminiscent of results obtained during analysis of the gene for transforming growth factor β3 (TGF-β3) in selected human breast cancer cell lines (Arrick et al. 1994). In that study, the shorter TGF-β3 transcript initiated downstream of 9 of 11 uORFs, and polysome analysis of the shorter transcript indicated that translation was enhanced relative to that of the transcript with the longer leader.

The potential significance of uORFs relates to the well documented observation that downstream AUG(s) are much less efficiently translated when uORFs are present in good context. This phenomenon is explained by the ribosomal scanning hypothesis, which asserts that binding of the 40 S ribosomal subunit occurs primarily at the 7mG cap and predicts that the first AUG codon in a good context is generally the one that initiates translation (Kozak 1991, 2002). Thus, uORFs beginning with AUGs in a poor context are not expected to appreciably affect the translation of a downstream polypeptide whose AUG is located in a favorable context (leaky scanning model). On the other hand, uORFs with AUGs in a good context are predicted to compete with downstream AUGs, which are proposed to be translated by the less efficient mechanism of ribosome reinitiation (Kozak 1991, 2002). Although it has been shown in some cases that the peptide encoded by an uORF has a direct and negative effect on the efficiency of translation of the downstream polypeptide (reviewed in Morris and Geballe 2000), other studies have shown that the *fact* of translation of the uORF, not its amino acid composition, is responsible for repression of translation from downstream ORFs (Damiani and Wessler 1993; Wang and Wessler 1998).

Only a small proportion (< 10%) of eukaryotic mRNAs appear to contain upstream AUG codons. However, these mRNAs seem to represent genes with essential regulatory functions (Kozak 1987), and it has been argued that, in cases where uAUGs are retained despite production of multiple leader sequences, there is a compelling reason to down-regulate translation, as mutations that release translational constraints on these genes are often associated with disease (Kozak 2002). Genes that can by-pass the reduction in translation efficiency imposed by the presence of uORFs usually do so by alternative splicing (reviewed in Kozak 2002), the use of an internal promoter to generate a downstream, monocistronic mRNA (see, for example, Han and Zhang 2003) or, for cellular mRNAs, more rarely by utilizing internal ribosomal entry sites (see, for example, Chappell and Mauro 2003).

Based on these observations and what is currently known about eukaryotic translational control, it appears that *inrpk1* is a member of that rare class of translationally controlled genes by virtue of its relatively long 5′ leader and the presence of multiple uAUGs. Both AUG_{590} and AUG_{668} are identical to AUG_{1144} at positions −3 and +4 (A in AUG is +1) and are good matches to consensus plant translation start sites (Fig. 5B). In contrast, the AUGs at nt 841, 852, 918 and 945 are poorer matches in comparison. This suggests that AUG_{590}, AUG_{668} and AUG_{1144} would be translated with essentially equivalent efficiencies, although it is difficult to predict how efficiently the others would be translated, if at all. Thus, the presence of AUG_{590} in transcripts originating from nt 561 (a small percentage of transcripts in nearly all the tissues examined) and of AUG_{668} in transcripts beginning at nt 590 (primarily

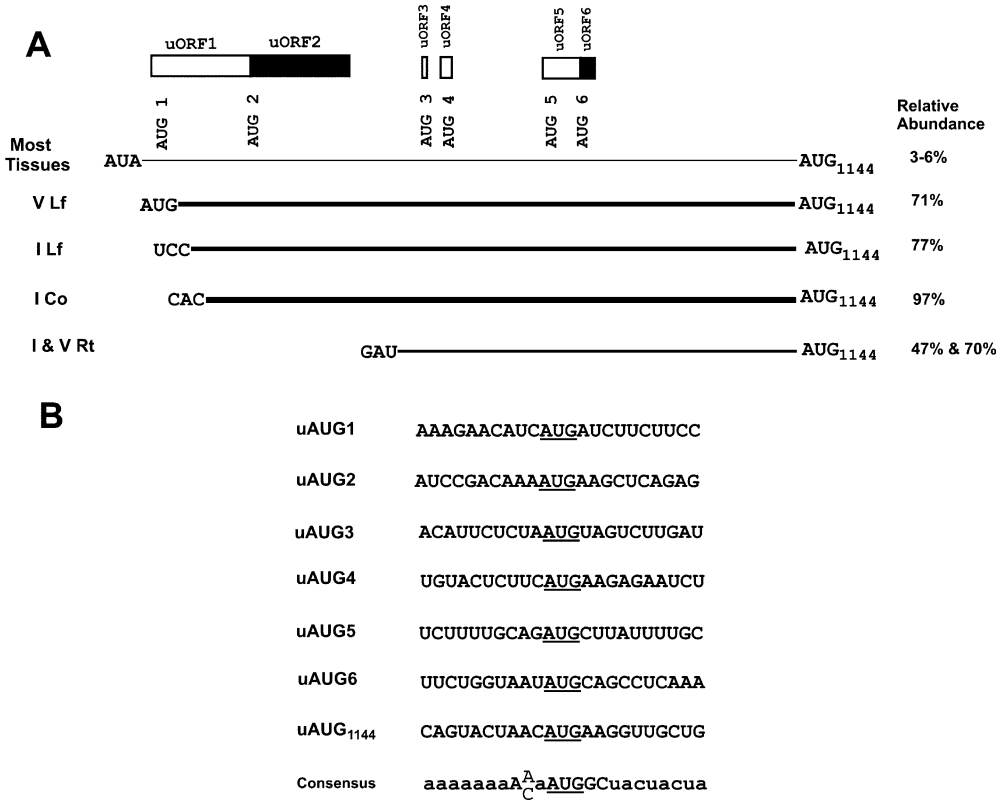

Fig. 5A, B Location and context of uORFs in the promoter of *inrpk1*. **A** Schematic representation of major and minor transcripts initiating from the promoter of *inrpk1* in various tissues from induced or vegetative plants. The positions of the uAUGs and uORFs are indicated. The *boxes* represent uORFs; the black boxes indicate uORFs that are in frame with the immediately preceding uORF. The relative abundance (percentage of clones) of each transcript is represented by the thickness of the line and indicated to the *right*. I, induced; V, vegetative; Lf, first leaf; Co, cotyledon; Rt, root. **B** Comparison of the context of the uAUGs with the plant dicot consensus sequence

leaves), nt 600 (induced cotyledons), nt 605 (induced leaves), and nt 643 (vegetative leaves) suggests that expression of INRPK1 is substantially reduced or even absent in these tissues, since these AUG codons would be encountered before AUG_{1144} in such transcripts. In contrast, the majority of mRNAs expressed in the roots originate at nt 716 or beyond, therefore bypassing the two uAUGs in good context, but retaining the four remaining uAUGs. Since no evidence of splicing was found for any of the samples examined and since previous 5′ RACE analysis of the region immediately upstream of AUG_{1144} failed to detect any transcripts initiating from this region (data not shown), i.e. evidence for a cryptic promoter was not found, it is likely that the four uORFs are retained in RNAs expressed in these tissues. Therefore in roots, transcripts could initiate translation at AUG_{1144} via a leaky scanning-type mechanism that skips the four uAUGs in poor context or possibly via reinitiation if the uORF5 peptide is small enough (Kozak 2001). In either case, translation of *inrpk1* is predicted to be less efficient due to the upstream AUGs. As a result, expression of *inrpk1* in roots would be regulated at three levels: transcription initiation, mRNA processing (Bassett et al. 2000) and translation initiation. Experiments are currently being designed to assess whether or not the longer transcripts observed in aerial tissues can be translated, and whether or not translation of *inrpk1* mRNA in roots is controlled primarily by the scanning mechanism or by some other process.

Acknowledgements The authors acknowledge that all work presented in this paper was carried out in compliance with current regulations guiding recombinant DNA experimentation. The authors would like to thank Dr. Tim Artlip for assistance with designing the gene-specific primer used in the RLM-5′RACE experiment. We would also like to acknowledge the expert technical assistance of Jami Young. As always, we acknowledge our debt to El Elyon.

References

Akiyama Y (1995) TFSEARCH: searching transcription factor binding sites. Available at http://www.cbrc.jp/research/db/TFSEARCH.html

Arrick BA, Grendell RL, Griffin LA (1994) Enhanced translational efficiency of a novel transforming growth factor 3 mRNA in human breast cancer cells. Mol Cell Biol 14:619–628

Bassett CL, Nickerson ML, Cohen RA, Rajeevan MS (2000) Alternative transcript initiation and novel post-transcriptional processing of a leucine-rich repeat receptor-like protein kinase gene that responds to short-day photoperiodic floral induction in morning glory (*Ipomoea nil*). Plant Mol Biol 43:43–58

Bucher P (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol 212:563–578

Burke TW, Kadonaga JT (1996) Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev 10:711–724

Burley SK, Kamada K (2002) Transcription factor complexes. Curr Opin Struct Biol 12:225–230

Chappell SA, Mauro VP (2003) The internal ribosome entry site (IRES) contained within the RNA-binding motif protein 3 (Rbm3) mRNA is composed of functionally distinct elements. J Biol Chem 278:33793–33800

Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 Angstrom resolution. Science 292:1863–1876

Damiani RD Jr, Wessler SR (1993) An upstream open reading frame represses expression of *Lc*, a member of the *R*/*B* family of maize transcriptional activators. Proc Natl Acad Sci USA 90:8244–8248

Fra AM, Pasqualetto E, Mancini M, Sitia R (2000) Genomic organization and transcriptional analysis of the human genes coding for caveolin-1 and caveolin-2. Gene 243:75–83

Gidoni D, Brosio P, Bond-Nutter D, Bedbrook J, Dunsmuir P (1989) Novel *cis* -acting elements in Petunia *cab* gene promoters. Mol Gen Genet 215:337–344

Haley J, Bogorad L (1990) Alternative promoters are used for genes within maize chloroplast polycistronic transcription units. Plant Cell 2:323–333

Han B, Zhang JT (2002) Regulation of gene expression by internal ribosome entry sites or cryptic promoters: the eIF4G story. Mol Cell Biol 22:7372–7384

Hochheimer A, Tjian R (2003) Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev 17:1309–1320

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res 15:6643–6651

Joshi CP, Zhou H, Huang X, Chiang VL (1997) Context sequences of translation initiation codon in plants. Plant Mol Biol 35:993–1001

Kondrakhin YV, Kel AE, Kolchanov NA, Romashchenko AG, Milanesi L (1995) Eukaryotic promoter recognition by binding sites for transcription factors. Comp Appl Biosci 11:477–488

Kozak M (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8148

Kozak M (1991) An analysis of vertebrate mRNA sequences: intimations of translational control. J Cell Biol 115:887–903

Kozak M (2001) Constraints on reinitiation of translation in mammals. Nucleic Acids Res 29:5226–5232

Kozak M (2002) Pushing the limits of the scanning mechanism for intiation of translation. Gene 299:1–34

Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev 12:34–44

Lee SW, Heinz R, Robb J, Nazar RN (1994) Differential utilization of alternate initiation sites in a plant defense gene responding to environmental stimuli. Eur J Biochem 226:109–114

Lewin B (2000) Genes VII. Oxford University Press. New York, NY

Luo M, Orsi R, Patrucco E, Pancaldi S, Cella R (1997) Multiple transcription start sites of the carrot dihydrofolate reductase-thymidylate synthase gene, and sub-cellular localization of the bifunctional protein. Plant Mol Biol 33:709–722

Lupold DS, Caoile AGFS, Stern DB (1999) The maize mitochondrial *cox2* gene has five promoters in two genomic regions, including a complex promoter consisting of seven overlapping units. J Biol Chem 274:3897–3903

Mauro VP, Nguyen T, Katinakis P, Verma DPS (1985) Primary structure of the soybean nodulin-23 gene and potential regulatory elements in the 5′-flanking regions of nodulin and leghemoglobin genes**.** Nucleic Acids Res 13:239–249

Meshi T, Iwabuchi M (1995) Plant transcription factors. Plant Cell Physiol 36:1405–1420

Michelet B, Lukaszewicz M, Dupriez V, Boutry M (1994) A plant plasma membrane proton-ATPase gene is regulated by development and environment and shows signs of a translational regulation. Plant Cell 6:1375–1389

Milanesi L, Muselli M, Arrigo P (1996) Hamming Clustering method for signal prediction in 5′ and 3′ regions of eukaryotic genes. Comp Appl Biosci 12:399–404

Mireau H, Lancelin D, Small ID (1996) The same Arabidopsis gene encodes both cytosolic and mitochondrial alanyl-tRNA synthetases. Plant Cell 8:1027–1039

Morris DR, Geballe AP (2000) Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol 20:8635–8642

Navarro-Avino JP, Hentzen AE, Bennett AB (1999) Alternative transcription initiation sites generate two *LCA1* Ca^{2+}-ATPase mRNA transcripts in tomato roots. Plant Mol Biol 40:133–140

Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, Burley SK (1999) TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev 13:3217–3230.

Pan S, Czarnecka-Verner E, Gurley WB (2000) Role of TATA binding protein-transcription factor IIB interaction in supporting basal and activated transcription in plant cells. Plant Cell 12:125–135

Prestridge DS (1991) SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. Comp Appl Biosci 7:203–206

Procissi A, Piazza P, Tonelli C (2002) A maize *r1* gene is regulated post-transcriptionally by differential splicing of its leader. Plant Molec Biol 49:239–248

Quandt K, Frech K, Karas H, Wingender E, Werner T (1995) MatInd and MatInspector—new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res 23:4878–4884

Reese MG, Harris NL, Eeckman FH (1995) Large scale sequencing specific neural networks for promoter and splice site recognition. In: Hunter L, Klein T (eds) Proceedings of the 1996 Pacific Symposium on Biocomputing, World Scientific Press, New Jersey, pp 776

Schug J, Overton GC (1997) TESS: Transcription Element Search Software on the WWW. Technical Report CBIL-TR-1997-1001-v0.0. Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania

henk T (1981) Transcriptional control regions: nucleotide sequence requirements for initiation by RNA polymerase II and III. Curr Topics Microbiol Immunol 93:25–46

Sheen J (1991) Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. Plant Cell 3:225–245

Short KC, Torrey JG (1972) Cytokinins in seedling roots of pea. Plant Physiol 49:155–160

Smale ST, Baltimore D (1989) The "initiator" as a transcription control element. Cell 57:103–113

Tamaoki M, Tsugawa H, Minami E, Kayano T, Yamamoto N, Kano-Murakami Y, Matsuoka M (1995) Alternative RNA products from a rice homeobox gene. Plant J 7:927–938

Walker JC, Zhang R (1990) Relationship of a putative receptor protein kinase from maize to the S-locus glycoproteins of *Brassica*. Nature 345:743–746

Wang L, Wessler SR (1998) Inefficient reinitiation is responsible for upstream open reading frame-mediated translational repression of the maize *R* gene. Plant Cell 10:1733–1745

Wang Y, Jensen RC, Stumph WE (1996) Role of TATA box sequence and orientation in determining RNA polymerase II/III transcription specificity. Nucleic Acids Res 24:3100–3106

Wassarman DA, Sauer F (2001). $TAF_{II}250$: a transcription toolbox. J Cell Sci 114:2895–2902

Zhang Y, Niu Z, Cohen AJ, Nah H-D, Adams SL (1997) The chick type III collagen gene contains two promoters that are preferentially expressed in different cell types and are separated by over 20 kb of DNA containing 23 exons. Nucleic Acids Res 25:2470–2477